Monte Carlo Methods and Appl., Vol. 11, No. 1, pp. 1 – 9 (2005)

On global sensitivity analysis of quasi-Monte Carlo algorithms

I.M.Sobol' [1] **and S.S.Kucherenko** [2]

[1] Institute for Mathematical Modelling
of the Russian Academy of Sciences,
4 Miusskaya Square, Moscow 125047, Russia

E-mail: hq@imamod.ru

[2] Imperial College London, SW7 2AZ, UK

E-mail: s.kucherenko@imperial.ac.uk

Abstracts — A Quasi-Monte Carlo algorithm corresponding to its Monte Carlo counterpart is not necessarily equivalent. Even in the case when their constructive dimensions are equal and the same quasi-random points are used, the efficiencies of these algorithms may differ. Global sensitivity analysis provides an insight into this situation. As a model problem two well-known approximations of a Wiener integral are considered: the standard one and the Brownian bridge. The advantage of the Brownian bridge is confirmed.

Keywords: Monte Carlo, Quasi-Monte Carlo, Global sensitivity analysis, Brownian bridge.

1 Introduction

Consider a Monte Carlo method for estimation of the mathematical expectation $I = E\eta$ of a random variable η:

$$\theta_N = \frac{1}{N}\sum_{k=1}^{N} \eta_k, \tag{1}$$

where η_k are independent realizations of η. Under the law of large numbers, $\theta_N \xrightarrow{P} I$ as $N \to \infty$. If the variance $D\eta$ is finite, the rate of convergence is $1/\sqrt{N}$. For a complete definition of the algorithm, it is necessary to provide a formula for modeling η:

$$\eta = g(\gamma_1, \gamma_2, ...) \tag{2}$$

where $\gamma_1, \gamma_2, ...$ are standard random numbers. The convergence rate of the algorithm (1), (2) does not depend on the choice of the function g.

For the quasi-Monte Carlo method corresponding to the algorithm (1), (2) the situation with the convergence rate is different. An essential role is played by the so-called constructive dimension (c.d.) of the algorithm. By definition, the c.d. of the algorithm (1), (2) is the maximum number of random numbers needed for the calculation of one value of η. In other words, c.d. $= n$ means that

$$\eta = g(\gamma_1, ..., \gamma_n). \tag{3}$$

The quasi-Monte Carlo algorithm corresponding to (1), (3) is

$$I_N = \frac{1}{N}\sum_{k=1}^{N} g(\vec{\mathrm{Q}}_k), \tag{4}$$

where $\{\vec{\mathrm{Q}}_k\}$ is a sequence of nonrandom points uniformly distributed in the unit hypercube, $\vec{\mathrm{Q}}_k = (q_k^1, ..., q_k^n)$. The Cartesian coordinates of the point $\vec{\mathrm{Q}}_k$ are used instead of random numbers for the realization of the k-th test.

For the best known sequences $\{\vec{\mathrm{Q}}_k\}$ and good enough functions g the estimate for the rate of convergence $I_N \to I$ is known to be $\mathrm{O}(\ln^n N)/N$. Apparently, the smaller n, the better the estimate. However in practice at $n > 1$, the law $\ln^n N/N$ is not observed (at least for $N < 10^7$) but it appears to be approximately $N^{-\alpha}$, $0 < \alpha \leq 1$.

In this paper we consider the Wiener path integral

$$I = \int_C F[x(t)]d_W x, \tag{5}$$

where $C$is the space of all functions $x(t)$ continuous in the interval $0 \leq t \leq T$ with the boundary condition $x(0) = x_0$. This integral is often encountered in probability theory, quantum physics, financial mathematics, etc.

The integral (5) can be regarded as an expectation with respect to the Wiener measure on C, so that $I = E(F[\xi(t)])$. Here $\xi(t)$ is a random Wiener process (also called a Brownian motion). Hence it is easy to apply a Monte Carlo approach which in this case would consist of constructing many random paths $\xi(t)$, evaluating $F[\xi(t)]$ and averaging the result. In addition the approximation algorithms for constructing $\xi(t)$ are required. We consider two such algorithms called discretization algorithms. The first one which is known as the Standard discretization follows directly from the definition of $\xi(t)$. The second or alternative discretization algorithm is based on the use of conditional distributions introduced by P. Levy [3]. Both algorithms were described in [5], where the functional

$$F[x(t)] = \int_0^T x^2(t)dt \tag{6}$$

was used as a test problem. In 1996, the alternative algorithm was reintrodused in [4]. It became known as the Brownian bridge. Both algorithms have the same variance, hence their Monte Carlo accuracies are equal. But as was shown in [4] the Brownian bridge is much more efficient in quasi-Monte Carlo.

A new concept called the "effective dimension" of an algorithm was introduced in [1] and [4]. The "effective dimension" may be smaller than the c.d. In our example it is smaller for the Brownian bridge than for the Standard algorithm, and this explains the superiority of the Brownian bridge. However, the proposed quantitative measure of the quality of quasi-Monte Carlo methods is rather complex and it depends on an arbitrary confidence level. Our approach is less general but easier for practical analysis.

2 Discretization of the Wiener process

We assume that the diffusion constant in the definition of Wiener's measure is $1/2$ and that the boundary value $\xi(0) = \xi_0$ is fixed. The interval $0 \le t \le T$ is divided into n equal parts. Random values of the process at the moments $t_i = (i/n)T, 1 \le i \le n$ are sampled using independent normal $N(0;1)$ variables ζ_i (we recall that $E\zeta_i = 0$, $E\zeta_i^2 = 1$, $Var(\zeta_i^2) = 2$). Adjacent points $(t_i, \xi(t_i))$ in the (t,x) plane are connected by straight lines, thus a continuous path $\xi(t)$ is replaced with a polygonal approximation $\xi_n(t)$.

The Standard algorithm is defined by the relation:

$$\xi(t_{i+1}) = \xi(t_i) + \sqrt{T/n}\zeta_{i+1}, 0 \le i \le n-1 \ . \tag{7}$$

The Brownian bridge algorithm is defined for $n = 2^p$ with an integer $p > 0$ by formulas:

$$\begin{aligned}
&\xi(T) = \xi_0 + \sqrt{T}\zeta_1, \\
&\xi(T/2) = \tfrac{1}{2}(\xi(T) + \xi_0) + \tfrac{1}{2}\sqrt{T}\zeta_2, \\
&\xi(T/4) = \tfrac{1}{2}(\xi(T/2) + \xi_0) + \tfrac{1}{2}\sqrt{T/2}\zeta_3, \\
&\xi(3T/4) = \tfrac{1}{2}(\xi(T/2) + \xi(T)) + \tfrac{1}{2}\sqrt{T/2}\zeta_4, \\
&\vdots \\
&\xi((n-1)T/n) = \tfrac{1}{2}(\xi((n-2)T/n) + \xi(T)) + \tfrac{1}{2}\sqrt{2T/n}\zeta_n.
\end{aligned} \tag{8}$$

The probability distributions of the paths $\xi_n(t)$ for both discretizations are the same, hence the variances $Var(F_n)$, where $F_n = F[\xi_n(t)]$ are equal.

Consider now the functional

$$F[\xi(t)] = \int_0^T \xi^2(t)dt. \tag{9}$$

The integral of $\xi_n^2(t)$ can be evaluated analytically [5,6].

Irrespective of the type of the approximation for $\xi_n(t)$, the expression for F_n has the general form

$$F_n = \sum_i a_i\zeta_i^2 + \sum_{i<j} a_{ij}\zeta_i\zeta_j. \tag{10}$$

Here a_i and a_{ij} are coefficients depending on n and on the type of approximation.

For numerical experiments we use $T = 1$ and $\xi_0 = 0$ so that $I = \frac{1}{2}$ and $D = \frac{1}{3}$. The coefficients a_i at $n = 4$ are

$$a_1 = \frac{10}{48}, a_2 = \frac{7}{48}, a_3 = \frac{4}{48}, a_4 = \frac{1}{48}$$

for the Standard algorithm, and

$$a_1 = \frac{16}{48}, a_2 = \frac{4}{48}, a_3 = \frac{1}{48}, a_4 = \frac{1}{48}$$

for the Brownian bridge. In general, a_i for the Standard algorithm decrease linearly with index i, and for the Brownian bridge they decrease much faster. Hence, there are fewer important square terms in (10) for the Brownian bridge and one can suggest that it is more

efficient. However, the influence of the products $\zeta_i\zeta_j$ (their number rapidly increases with n) remains unclear. Therefore we consider sensitivity indices S_i rather than coefficients a_i.

Remark. In (10) $\zeta_i = G(\gamma_i)$, where $G(y)$ is the inverse normal cumulative distribution function.

3 Global sensitivity indices

Global sensitivity indices were introduced in 1990 by I.M.Sobol' [7]. A more recent presentation can be found in [8]. Let $x = (x_1, x_2, ..., x_n)$ be a point in the n-dimensional unit hypercube. Consider a square integrable function $f(x)$ defined in the hypercube. Denote by f_0 its mean value

$$f_0 = \int_0^1 ... \int_0^1 f(x) dx_1 ... dx_n. \tag{11}$$

Definition. The representation

$$f(x) = f_0 + \sum_{s=1}^{n} \sum_{i_1 < ... < i_s} f_{i_1 ... i_s}(x_{i_1}, ..., x_{i_s}) \tag{12}$$

is called ANOVA-decomposition if

$$\int_0^1 f_{i_1 ... i_s} dx_{i_k} = 0 \qquad \text{for} \qquad 1 \le k \le s. \tag{13}$$

Here $1 \le \quad i_1 < ... < i_s \le \quad n$. Conditions (11) and (13) define uniquely all terms in (12).

Formula (12) can be written as

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + ... + f_{12...n}(x_1, x_2, ..., x_n).$$

Now assume that $x_{1_1}, ..., x_n$ are independent random variables with distribution functions $F_1(x_1), ..., F_n(x_n)$ and $f(x_{1_1}, ..., x_n)$ is a random variable with a finite variance

$$D = \mathrm{Var}(f).$$

In this case the definition of ANOVA is practically the same as before: instead of (11) we write an expectation $f_0 = E(f(x))$ or explicitly

$$f_0 = \int ... \int f(x) dF_1(x_1) ... dF_n(x_n) \ ,$$

and the condition analogous to (13) is written as

$$\int f_{i_1 ... i_s} dF_{i_k}(x_{i_k}) = 0 \quad \text{for} \qquad 1 \le k \le s.$$

Global sensitivity indices $S_{i_1 ... i_s}$ are defined as ratios

$$S_{i_1 \dots i_s} = D_{i_1 \dots i_s}/D,$$

where $D_{i_1 \dots i_s} = \mathrm{Var}(f_{i_1 \dots i_s})$.

¿From (12) one can see that the sum of all sensitivity indices is equal to 1:

$$\sum_{s=1}^{n} \sum_{i_1 < \dots < i_s} S_{i_1 \dots i_s} = 1.$$

Clearly, $f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) \equiv 0$ if and only if the index $S_{i_1 \dots i_s} = 0$. The most important property of the ANOVA-decomposition is the possibility to estimate the indices numerically directly from the values of $f(x_1, \dots, x_n)$. Global sensitivity indices are widely used as a tool for analyzing complex functions [7, 8]. In particular,

$$\sum_{i=1}^{n} S_i = 1$$

if and only if

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i).$$

The index S_i shows the direct influence of x_i on the output f. The total influence of x_i on f can be characterized by S_i^{tot}. S_i^{tot} is defined as the sum of all $S_{i_1 \dots i_s}$, such that one of the indices $i_1, \dots, i_s$ is equal to i, see [8] for details.

The "effective dimension" of $f(x)$ was defined in [1] by using truncated ANOVA decompositions.

Remarks. Functional relations that include random variables are true with probability 1. The term ANOVA comes from Analysis of Variances. In the following example the sensitivity indices can be computed analytically.

4 Global sensitivity analysis of F_n

¿From (10) we find the expectation

$$I_n = E(F_n) = \sum_i a_i \tag{14}$$

and the ANOVA decomposition

$$F_n = I_n + \sum_i a_i(\zeta_i^2 - 1) + \sum_{i<j} a_{ij}\zeta_i\zeta_j. \tag{15}$$

Thus F_n consists of one-dimensional and two-dimensional terms only. ¿From (15) it is also easy to find the variance

$$D(F_n) = 2\sum_i a_i^2 + \sum_{i<j} a_{ij}^2. \tag{16}$$

Therefore the first order indices are

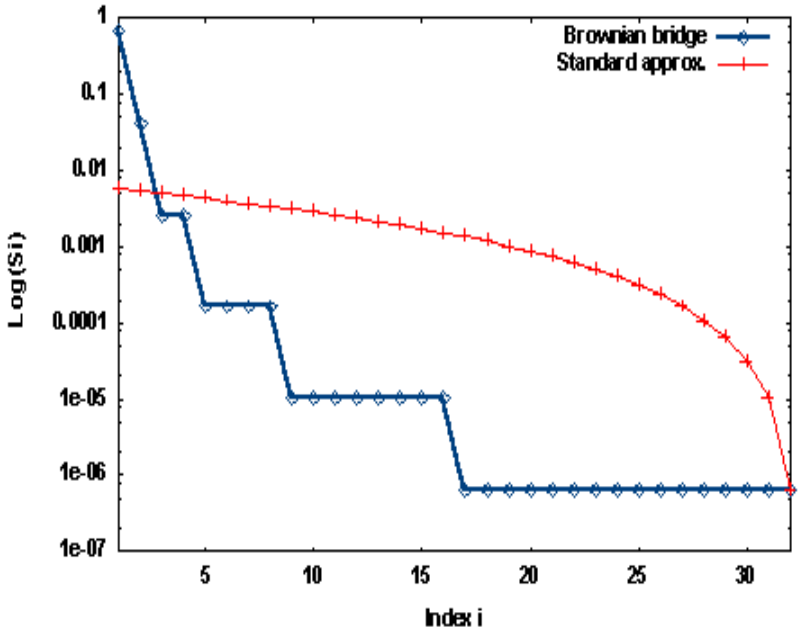

Figure 1. Log_{10} of first order sensitivity indices S_i versus index number i, n=32.

Table 1.

n	I_n	$D(F_n)$	$\sum_i S_i$ Stand.	$\sum_i S_i$ BB
4	0.458	0.323	0.4367	0.7207
8	0.479	0.331	0.2361	0.7214
16	0.489	0.332	0.1222	0.7215
32	0.495	0.333	0.0612	0.7215

$$S_i = 2a_i^2/D(F_n) \tag{17}$$

and the second order indices are

$$S_{ij} = a_{ij}^2/D(F_n).$$

We recall that the variances $D(F_n)$ are the same for both algorithms, so they are equivalent as far as the Monte Carlo method is concerned.

At first glance, one should not expect too much information from indices S_i: they are simply proportional to a_i^2. However, we recall that the sum of all $S_{i_1 \ldots i_s}$ is always 1 and the knowledge of the sum of all S_i is rather useful.

The values of S_i at n=32 are shown in Fig.1.

Table 1 contains the values of $\sum_{i=1}^{n} S_i$ for both algorithms at different n.

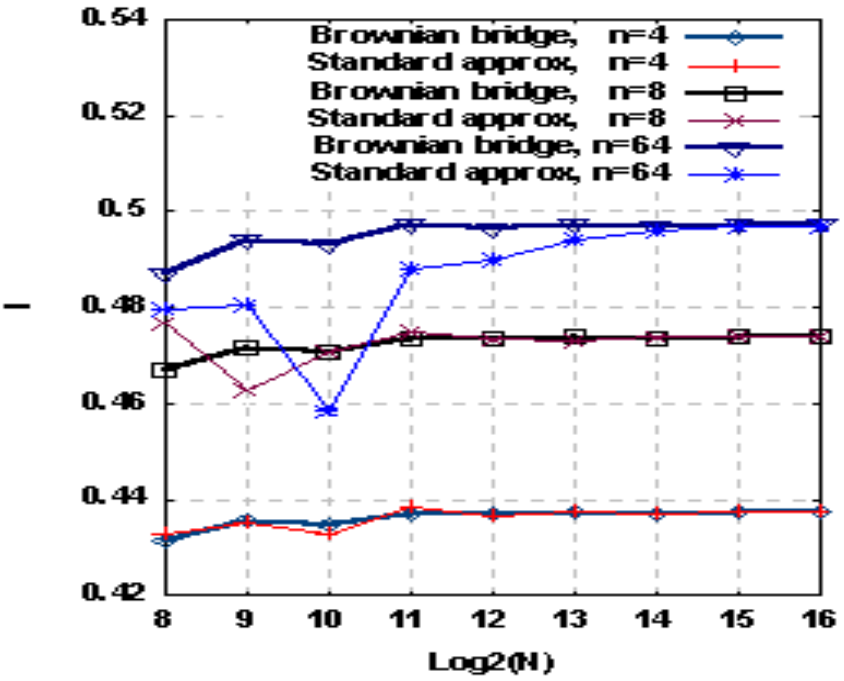

Figure 2. $I_{n,N}$ versus $\mathrm{Log}_2(N)$ at n=4, 8, 64.

Clearly, for the Brownian bridge the main contribution to F_n comes from one-dimensional terms (approximately 72%), while for the Standard algorithm at $n > 4$ the main contribution to F_n is defined by two-dimensional terms.

As a rule, in quasi-Monte Carlo one-dimensional integrals are evaluated with greater accuracy than integrals of higher dimension. Therefore in quasi-Monte Carlo the Brownian bridge is superior to the Standard algorithm.

In our numerical experiments LP_τ - sequences (or (t,s) - sequences in base 2) often called Sobol sequences were used. All finite-dimensional projections of such sequences are also LP_τ - sequences. Even more, all one-dimensional projections are in a sence optimal: the values of τ are the least possible: τ=0. On the contrary, for two-dimensional projections onto the (x_i, x_j)-plane the values of τ slowly increase when i or j increase [2,9].

5 Numerical results

Fig. 2 shows results of numerical evaluation of I given by (5),(6) at different n as functions of N. $I_{n,N}$ converges to I_n as $N \to \infty$. Both algorithms were implemented with the same points of Sobol' sequence. There is practically no difference in the performance at $n = 4$ but for higher n the Brownian bridge results (solid lines) converge considerably faster.

Fig. 3 shows integration errors ε versus N, at n=64. To reduce the scatter in error values the errors were averaged over K=50 independent runs [4]:

$$\varepsilon = \left(\frac{1}{K} \sum_{k=1}^{K} (I_n - I_{n,N}^k)^2 \right)^{1/2} .$$

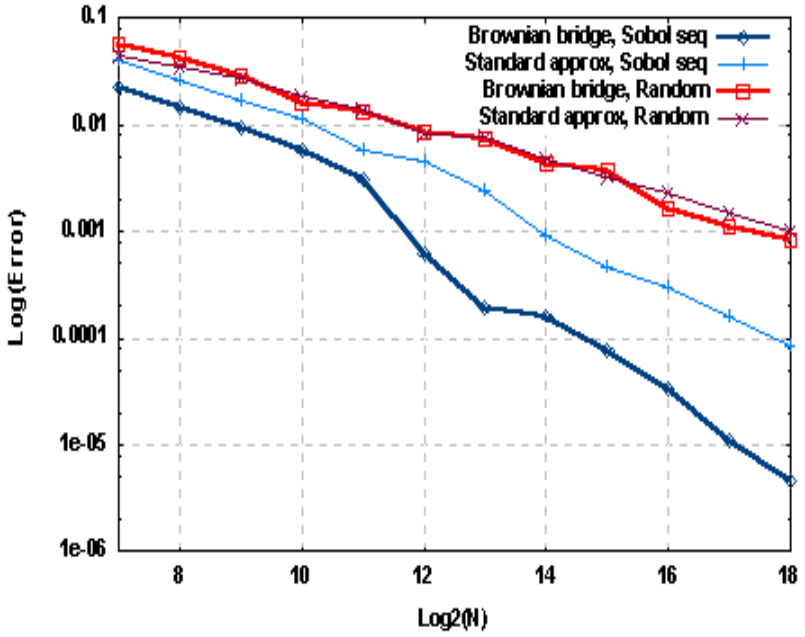

Figure 3. Log-log plot of the root mean square errors ε versus N at n=64.

In the Monte Carlo computations, the runs use different pseudo-random numbers. For the quasi- Monte Carlo computations, different sections of Sobol' sequence were used.

In full agreement with the discussions above, in the case of pseudo-random numbers both algorithms produce similar errors. However the quasi-Monte Carlo errors are rather different: the errors of the Brownian bridge are much lower. From the last five points of each curve the convergence rate can be estimated: it is approximately $O(N^{-1/2})$ for the Monte Carlo results, while for both quasi-Monte Carlo curves it is $O(N^{-1})$.

6 Conclusions

Global sensitivity analysis can be successfully applied to quasi-Monte Carlo algorithms. The first order sensitivity indices used in this paper are less general than the "effective dimension" approach but much easier for computing.

New evidence for superiority of the Brownian bridge is given.

Acknowledgments

One of the authors (SK) gratefully acknowledges the financial support of the United Kingdom's Engineering and Physical Sciences Research Council (EPSRC) under Platform Grant GR/N08636. The other author (IS) was supported by the RFBR Grant N 04-01-00-367.

References

[1] R.E. Caflisch, W. Morokoff, A.B. Owen. Valuation of Mortgage Backed Securities using Brownian Bridges to Reduce Effective Dimension. The Journal of Computational Finance, 1997, V. 1, N 1, 27-46.

[2] P. Jaeckel. Monte Carlo Methods in Finance. J.Wiley, London, 2002.

[3] P. Lévy Processus stochastiques et mouvement brownien, Paris : Gauthier-Villars, 1965.

[4] B. Moskowitz, R.E. Caflisch, Smoothness and Dimension Reduction in Quasi-Monte-Carlo Methods, Mathl. Comput. Modelling, 1996, V. 23, N 8/9, 37-54.

[5] I.M.Sobol'. Computation of definite integrals, in "The Monte Carlo Method", ed. Yu. A. Shreider, Pergamon Press, 1966, translated from "The method of statistical trials", Fizmatgiz, Moscow, 1962 (in Russian).

[6] I.M.Sobol'. Numerical methods of Monte Carlo. Moscow, 1973 (in Russian).

[7] I.M.Sobol'. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie, 1990, V. 2, N 1, 112-118 (in Russian), English translation in: Mathematical Modeling and Computational Experiment, 1993, V. 1, 407-414.

[8] I.M.Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and computers in simulation, 2001, V. 55, 271-280.

[9] I.M.Sobol', V.I.Turchaninov, Yu.L.Levitan, B.V.Shukhman. Quasirandom sequence generators. Keldysh Inst. Appl. Math, Moscow, 1992.